

Geostatistical inference of hydrogeological parameters with a large number of unknowns

Doctoral Thesis**Author(s):**

Li, Wei

Publication date:

2008

Permanent link:

<https://doi.org/10.3929/ethz-a-005724349>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Diss. ETH No. 17865

Geostatistical Inference of Hydrogeological Parameters with a Large Number of Unknowns

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Science

presented by
WEI LI
M.Sc.

born November 28, 1977
citizen of HeBei, China

accepted on the recommendation of
Prof. Dr. Peter Reichert, examiner
PD Dr.-Ing. Olaf A. Cirpka, co-examiner
Prof. Dr.-Ing. Wolfgang Kinzelbach, co-examiner

2008

Abstract

Flow and transport models in porous media have become a very important tool in groundwater management. To implement these models, spatial distributions of hydrogeological parameters must be known. Due to scarcity of direct measurements, the parameters are mainly inferred from secondary dependent information such as head or concentration measurements. The inference of hydrogeological parameters from secondary information is referred to as inverse problem of groundwater modeling. Because of consistent assumptions about the spatial variability of porous media, inverse methods within the framework of geostatistics are well suited for estimating the spatial distributions of hydrogeological parameters in heterogeneous aquifers. In geostatistics, the parameter field is regarded as a spatial random variable.

Within the geostatistical inverse approaches, the Quasi-Linear Geostatistical Inverse Approach of Kitanidis [1995] with its modifications appears to be the most efficient, particularly with respect to required CPU time in quantifying the expected value of the parameters and their related estimation uncertainty for the cases with a large number of unknowns. Despite this advantage, a couple of difficulties of applying this inverse approach remain.

The inverse approach requires evaluation of cross-covariance matrices in the inference. Efficient spectral methods of computing these terms [Nowak et al., 2003] require regular structured grids and will fail when the domain is discretized by unstructured grids which may be important for practical applications. Explicitly computing these matrices is prohibitive for cases with a large number of unknowns. In this thesis, I parameterize the parameter field using a

truncated series of base functions efficiently derived from the covariance matrix by spectral methods. The parameterization of a spatial random field using truncated series is the application of the Karhunen-Loève(KL) expansion [Loève, 1977]. The reduced dimension in the parameterization can reduce the computational costs in estimating the parameter field. The base functions are continuous in space and do not distinguish discretization schemes.

I integrate the parameterization with the KL expansion into the inverse approach and implement this integrated inverse approach using structured and unstructured grids in synthetic test cases. I compare the inverse approach with the KL expansion and the inverse approach with the full covariance matrix in which the FFT method of accelerating evaluation of cross-covariance matrices can be applied. The computational effort, estimates, and estimation variances are investigated. Results show that I can obtain a reliable estimate of hydrogeological parameters using a limited number of truncated terms. If a high number of KL terms are needed, the inverse approach with the full covariance matrix outperforms the inverse approach with the KL expansion. However, for smooth covariance functions, particularly with large correlation lengths, a few KL terms are sufficient and the parameterization with the KL expansion becomes more efficient. This work has been published by Li and Cirpka [2006].

Applicability of the inverse approach in analyzing field data is investigated. In this thesis, two applications with different spatial dimensions, a two-dimensional and a three-dimensional inference, are conducted. In the field applications, the measurements of drawdown during pumping tests and discharge profiles of flowmeter tests in fully screened wells at the test site Krauthausen, Germany, are used. For these two estimations, I apply the inverse approach with full covariance matrix on regular grids and accelerate the evaluation of cross-covariance matrices using the FFT method.

In the two-dimensional analysis, I test the feasibility of the inverse approach given measurements of pumping tests in estimating the fields of transmissivity and storativity. To estimate the field of storativity, I have to apply transient data of drawdown. To avoid high correlation of transient data in a drawdown curve, I

apply temporal moments of drawdowns [Li et al., 2005] to extract the most important information of transient curves. The corresponding moment-generating equations are at steady state, which can be solved much more efficiently than the transient flow equations.

For the estimated transmissivity, the inverse approach provides estimates showing more spatial variability than the interpolated field of the values obtained by the conventional type-curve approach. This comes from the consistent assumption of the former approach on the structure of porous media. Structures are obtained in the regions where the pumping tests were conducted and uncertainties of the estimate are decreased in these regions. However, uncertainties are still high for the areas far away from well locations. Considering the estimate of storativity, both approaches obtain a high value of estimated variance, which is believed unrealistic, because the variability of all terms making up the storativity is small in space so that the distribution of storativity is presumed to show small variance. I believe that this is an effect of aliasing. The estimated distribution of transmissivity is smoother than the real field. The unresolved variability at small local scales has a large effect on the simulated first temporal moments, representing the characteristic time of drawdown, than the zeroth temporal moments, representing final drawdown. Given a smooth estimate of transmissivity, the inverse approach attributes the derived variability in first temporal moments to the variability of storativity. In the estimation, I assumed a two-dimensional aquifer. The unresolved vertical variability may also be a particular cause for the unrealistic results. Data quality influences the estimation of parameter fields and geostatistical parameters. When the measurement error is large, I obtain increased values of estimated correlation lengths and decreased values of variances, which smooth the estimated fields of hydrogeological parameters. The two-dimensional study has been published by Li et al. [2007].

I conduct a three-dimensional estimation of hydrogeological parameters at the same site to test the performance of the inverse approach with a large number of unknowns. Due to lack of measurements supporting a three-dimensional estimation of specific storage coefficient, particularly regarding to the vertical

direction, the field of specific storage coefficient is not estimated. In the three-dimensional estimation, I jointly apply the data of pumping and flowmeter tests to obtain the field of hydraulic conductivity. Measurements of pumping tests reflect depth-averaged horizontal features of aquifers, whereas the flowmeter data contain the relative vertical distributions of hydraulic conductivity. I use final drawdown measurements of pumping tests and discharge profiles of the flowmeter tests as my data for the inversion. By considering discharge profiles, I avoid converting the relative hydraulic conductivities to absolute values making the integration of flowmeter data into the analysis of pumping tests consistent.

With about one million of unknowns, the inverse approach used roughly two days CPU time to finish the estimation of hydraulic conductivity on a standard personal computer. The results show that three-dimensional structures are obtained in the vicinity of the wells where flowmeter tests were conducted. In the region where only pumping test data exist, the estimated hydraulic conductivity becomes vertically uniform. The remaining uncertainty of the estimate decreases considerably near the wells for flowmeter tests. The three-dimensional study has been published by Li et al. [2008].

Overall, with the geostatistical inverse approach using a sufficient number of measurements with suitable support volumes, I can obtain reliable estimates of spatial distributions of hydrogeological parameters with reasonable computational effort. On unstructured grids, the inverse approach with the KL expansion [Li and Cirpka, 2006] appears to be the most efficient method. On structured grids, the method using the full covariance matrix can be applied to problems with up to millions of unknowns on a standard personal computer. In both methods, a good spatial distribution of the estimated fields is obtained mainly in the regions where aquifer tests are available. In order to identify all relevant structures in a formation, about one measurement per correlation length is required. However, traditional hydraulic investigation techniques such as pumping and flowmeter tests are costly because they require wells. Efficient data-acquisition techniques based on geophysical monitoring of hydraulic tests such as electrical resistivity tomography and a proper integration of these techniques into ground-

water inverse modeling seem necessary.

Zusammenfassung

Strömungs- und Transportmodelle sind wichtige Werkzeuge in der Bewirtschaftung von Grundwasserressourcen. Zu Ihrer Verwendung muss die räumliche Verteilung hydrogeologischer Parameter, namentlich der Durchlässigkeit und des Speicherkoeffizienten, bekannt sein. Weil direkte Messungen dieser Grössen selten sind, werden sie meistens aus Messungen abhängiger Grössen, wie dem Grundwasserspiegel oder der Konzentration eines gelösten Stoffes, abgeleitet. Die Ermittlung der hydrogeologischen Kenngrössen aus sekundärer Information wird als inverses Problem der Grundwassermodellierung bezeichnet. Inverse Methoden auf der Grundlage einer geostatistischen Beschreibung des Grundwasserleiters sind gut geeignet um die räumliche Verteilung hydrogeologischer Parameter in heterogenen Grundwasserleitern zu ermitteln, weil die geostatistische Charakterisierung eine konsistente Grundlage für die Beschreibung räumlich variabler Grössen darstellt. Konzeptionell beruht die Geostatistik auf der Annahme, dass das Parameterfeld als räumliche Zufallsvariable betrachtet werden kann.

Innerhalb der geostatistischen inversen Methoden ist der quasi-lineare geostatistische Ansatz von Kitanidis [1995] - mit verschiedenen Modifikationen - besonders günstig, vor allen Dingen in Bezug auf die rechnerische Effizienz in der Abschätzung des besten Schätzwertes und der damit verbundenen Schätzunsicherheit. Nichtsdestotrotz verbleiben Schwierigkeiten in der Anwendung der genannten Methode.

Der gewählte inverse Ansatz erfordert die Berechnung von Kreuz-Kovarianzmatrizen, die die Korrelation zwischen allen gemessenen Grössen und allen

diskretisierten Parametern beschreiben. Effiziente spektrale Methoden zur Berechnung dieser Matrizen [Nowak et al., 2003] erfordern die Diskretisierung des Gebietes mittels regelmässiger strukturierter Gitter und können nicht angewendet werden, wenn das Gebiet durch unstrukturierte Gitter diskretisiert wird, wie dies in vielen praktischen Anwendungen der Fall ist. Eine explizite Berechnung der Kreuz-Kovarianz-matrizen ist bei grossskaligen Anwendungen praktisch nicht möglich. In dieser Dissertationsschrift parametrisiere ich das räumlich kontinuierliche Parameterfeld mittels einer abgebrochenen Reihe von Basisfunktionen, die mit spektralen Methoden aus der Auto-Kovarianzfunktion abgeleitet werden können. Die Entwicklung ist als Karhunen-Loève-Entwicklung [Loève, 1977] bekannt. Durch Einbettung des Berechnungsgebietes in eine grössere Einheitszelle eines periodischen Gebietes können schnelle Fouriertransformationmethoden verwendet werden. Ausserdem ist die Form der Basisfunktionen analytisch vorgegeben. Der Abbruch der Entwicklungsreihe nach Berücksichtigung aller dominanten Terme kann zu einer Verringerung des Rechenaufwandes führen. Die Basisfunktionen sind kontinuierliche trigonometrische Funktionen und können auf beliebigen Gittern abgebildet werden.

Ich habe die Parametrisierung des Parameterfeldes mittels der Karhunen-Loève-Entwicklung in den genannten inversen Ansatz integriert und implementierte ein vollständiges inverses Modell für strukturierte und unstrukturierte Gitter. Ich führte Vergleichsrechnungen anhand synthetischer Testbeispiele durch. Im Vergleich zum inversen Ansatz unter Verwendung der vollständigen Kovarianzfunktion ist der Ansatz unter Anwendung der Karhunen-Loève-Entwicklung auf strukturierten Gittern etwas aufwändiger, zumindest wenn viele Karhunen-Loève-Terme berücksichtigt werden müssen. Im Fall glatter Kovarianzfunktionen mit grosser Korrelationslänge kann jedoch die Karhunen-Loève-Entwicklung nach vergleichsweise wenigen Gliedern abgebrochen werden, sodass die Entwicklung rechnerische Vorteile bietet. Auf unstrukturierten Gittern ist mir keine effizientere Methode zur Berechnung von Kreuzkovarianztermen bekannt. Diese Arbeiten wurden von Li and Cirpka [2006] veröffentlicht.

Die vorliegende Dissertationsschrift behandelt inverse Methoden zur Analyse

von Felddaten. Ich zeige zwei Feldanwendungen mit unterschiedlicher Dimensionalität. In diesen Anwendungen werden Pumpversuche und Flowmeter-Daten vom Testfeld Krauthausen des Forschungszentrums Jülich ausgewertet. Die Auswertung der Pumpversuche, die in vollständig verfilterten Brunnen durchgeführt wurden, erfolgte mit einem zweidimensionalen Modell. Bei Berücksichtigung der Flowmeterdaten wechselte ich zu einer dreidimensionalen Betrachtung. Im Testfeld wurden 24 Kleinpumpversuche unter Verwendung von 52 Beobachtungsrohren durchgeführt. Dies führte zu 179 Absenkungskurven. Flowmeter-Messungen lagen für 22 Beobachtungsrohre vor.

In der zweidimensionalen Anwendung versuchte ich aus instationären Pumpversuchsdaten die räumliche Verteilung der Transmissivität und des Speicherkoefizienten zu ermitteln. Da instationäre Daten zeitlich stark korreliert sind, charakterisierte ich die Absenkungskurven mittels ihrer zeitlichen Momente [Li et al., 2005]. Die zugehörigen momentengenerierenden Gleichungen entsprechen stationären Absenkungsgleichungen mit verteiltem Quell-/Senkenterm.

Die Absenkungskurven wurden zum Vergleich mittels des Theiss-Typkurvenverfahrens ausgewertet. Hierbei ergab sich, dass mit dem inversen Verfahren die räumliche Variabilität der Transmissivität besser aufgelöst werden konnte. Das Typkurvenverfahren beruht auf der Annahme eines homogenen Grundwasserleiters, was offensichtlich im Widerspruch zu Daten steht, die für unterschiedliche Kombinationen von Pump- und Beobachtungsbrunnen zu unterschiedlichen Werten führen. Die inverse Methode ermöglicht es, hydraulische Strukturen in der Nähe von Beobachtungsrohren zu identifizieren. In diesen Bereichen wird die Unsicherheit signifikant verringert. In grösserer Entfernung zu den Beobachtungsrohren verbleibt die Unsicherheit jedoch auf hohem Niveau. Die ermittelten räumlichen Felder des Speicherkoefizienten zeigen eine starke räumliche Variabilität. Diese Variabilität erscheint unrealistisch, weil die Kenngrößen, die zum Speicherkoefizienten beitragen (Porosität, Kompressibilität des Wassers und des Porenraums) nur geringfügig schwanken. Ich vermute, dass die unaufgelöste Variabilität der Durchlässigkeit zu einer systematischen Überschätzung der Variabilität des Speicherkoefizienten führt. Die Schätzung der geostatistis-

chen Kenngrössen hängt darüberhinaus stark von der Messgenauigkeit der Absenkung ab. Ein grosser Messfehler führt zu einer grossen geschätzten Korrelationslänge und einer kleinen Varianz und damit zu glatten Feldern der geschätzten hydrogeologischen Parameter. Die zweidimensionale Studie wurde von [Li et al., 2007] veröffentlicht.

Die dreidimensionale Anwendung beruht auf denselben Daten wie das zweidimensionale Beispiel. Zusätzlich verwendete ich die vorliegenden Flowmeterdaten. In dieser Anwendung schätzte ich lediglich die Durchlässigkeitsverteilung ab, weil keine Messungen vorlagen, die auf die dreidimensionale Verteilung des spezifischen Speicherkoeffizienten sensitiv wären. Die Flowmeterdaten geben ein Vertikalprofil der relativen Durchlässigkeit innerhalb des Profiles wieder. In meinem inversen Algorithmus verwende ich direkt die Durchflussprofile der Flowmetertests. Es ist also nicht notwendig, die Flowmeterdaten zunächst in absolute Durchlässigkeiten umzuwandeln. Letzteres wäre am Standort auch nicht möglich, weil tiefenintegrierte direkte Messungen der Durchlässigkeit fehlen.

Das Berechnungsgebiet ist in etwa eine Million Finite Elemente unterteilt. Für jedes Element schätzte ich die Durchlässigkeit ab. Dies erforderte etwa zwei Tage Rechenzeit auf einem Personal Computer. Die ermittelte dreidimensionale Durchlässigkeitsverteilung zeigt eine gute vertikale Auflösung in der Nähe von Beobachtungsrohren, in denen Flowmeter-Versuche durchgeführt wurden. In Gebieten, die von Flowmeter-Brunnen weit entfernt liegen, konnte keine vertikale Differenzierung erreicht werden. Die Schätzvarianz wird in der Nähe der Flowmeter-Brunnen stark reduziert. Die dreidimensionale Studie wurde von Li et al. [2008] veröffentlicht.

Zusammenfassend kann festgestellt werden, dass es möglich ist, mit dem geostatistischen inversen Ansatz zuverlässige Schätzungen zur Verteilung hydrogeologischer Kenngrössen vorzunehmen, sofern eine ausreichende Anzahl an Messungen mit angemessenem Messvolumen vorliegen. Der zugehörige Rechenaufwand ist auch für Probleme mit vielen Unbekannten erträglich. Auf unstrukturierten Gittern erscheint der Ansatz unter Verwendung der Karhunen-Loève-Entwicklung [Li and Cirpka, 2006] am effizientesten. Auf regelmässigen struk-

turierten Gittern können bis zu eine Million Parameter mit der Methode unter Verwendung der vollen Kovarianzmatrix auf einem Personal Computer ermittelt werden. Beide Methoden führen zu einer guten räumlichen Auflösung in der Nähe von Beobachtungspunkten. Um alle hydraulisch relevanten Strukturen im Untergrund zu identifizieren muss etwa ein Beobachtungspunkt je Korrelationslänge vorliegen. Hierin liegt eine der wesentlichen Beschränkungen in der Anwendung hydraulischer Versuche. Die Installation eines Beobachtungsrohres ist aufwändig und kostspielig, sodass an den meisten Feldstandorten keine ausreichende räumliche Auflösung der hydrogeologischen Parameterfelder durch die direkte Auswertung hydraulischer Versuche erreicht werden kann. Als effiziente Alternative zu direkten Beobachtungen könnten geophysikalische Messmethoden für das Monitoring hydraulischer Versuche eingesetzt werden. Allerdings erfordert dies die Entwicklung spezieller Methoden, um diese Messdaten in das inverse Schema für die Ermittlung hydrogeologischer Größen zu integrieren.